POLICY RESEARCH WORKING PAPER 1670

Assessing the Welfare Impacts of Public Spending

Dominique van de Walle

We must diversify and compare results from our methods of assessment, as well as broaden our definition of well-being, to see how public spending policies affect various facets of living standards.

The World Bank
Policy Research Department
Public Economics Division
October 1996

POLICY RESEARCH WORKING PAPER 1670

Summary findings

An important objective of public spending is to raise household living standards, particularly for the poor. But how can final impacts on this objective best be assessed? Evaluating a policy's impact requires assessing how different things would have been in its absence. But the counterfactual of no intervention is often tricky to quantify.

Van de Walle surveys the methods most often used to assess the welfare effects of public spending. In studying the current state of the art she identifies some limitations of current practices and draws implications for best practice in future work. The methods used to assess welfare impacts broadly fall into two groups: benefit incidence studies and behavioral approaches. Both have their strengths and weaknesses.

Benefit incidence studies ignore behavioral responses and second-round effects, and simply use the cost of provision as a proxy for benefits received. Behavioral approaches present quite different drawbacks, in attempting to represent individual benefits correctly.

A number of recent studies usefully combine both approaches.

It is still uncertain whether behaviorally consistent methods actually point to fundamentally different policy recommendations. What can be concluded is that we need to diversify and compare results from our evaluation methods and broaden our definition of well-being, to see how various facets of living standards are affected by public spending.

This paper — a product of the Public Economics Division, Policy Research Department — is part of a larger effort in the department to analyze public spending and poverty issues. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Cynthia Bernardo, room N10-055, telephone 202-473-1148, fax 202-522-1154, Internet address cbernardo@worldbank.org. October 1996. (31 pages)

Assessing the Welfare Impacts of Public Spending

Dominique van de Walle*

* The author is with the Public Economics Division in the Policy Research Department of the World Bank, Washington, D.C. She thanks Lionel Demery, Shanta Devarajan, Mark Gersovitz, David Sahn and Martin Ravallion for valuable comments.

Introduction

The role of government is to enhance social welfare. One way to achieve this is by spending on things of value to people which they cannot otherwise attain. Priority should be given to things that the private sector cannot or will not spend on, or that the private sector tends to under-provide. One of those things is equity. While most of us would prefer to see less inequality and poverty, individually we do not have much of an incentive to do something about it, since a large share of the benefits go to others. Thus, there can be an important case for public intervention to help improve distributional outcomes. Public spending is a potentially powerful instrument—and the main instrument available to governments in developing countries—for fighting poverty. But budgets are limited. Other things will be underprovided without public action. So governments face competing demands, and intense scrutiny of whether desirable distributional objectives are indeed being met by expenditure practices. Economists and policymakers therefore need to be able to evaluate the distributional impacts of public spending policies.

This paper aims to provide a critical, though selective, overview of the current state of practice in empirically assessing the welfare impacts of public spending. I assume that we are interested in assessing impacts on levels of some welfare indicator, and that greater weight is attached to people who are poorer in terms of that indicator. The assessment is thus concerned with inequality as well as with the average level of living. The aim is to give an overview of the main methods used, to discuss some of the key issues that have been raised about those methods and to see what all this means for policy.

Before tackling the question of how to measure the distributional effects of public spending, it is necessary to establish criteria by which performance will be assessed. Specifying the welfare objective—and why confusion about the welfare objective is often an obstacle to determining policy impacts—are the issues delved into in the paper's next section. This is followed by a discussion of common approaches to measuring distributional impacts. The limitations and strengths of "benefit incidence studies" and of "behavioral approaches" are explored with an emphasis on what they can and cannot tell us and how confident we can be in drawing policy implications. Recent studies which wed the two methods are also reviewed. The conclusion offers a brief summing up.

How Should We Judge Welfare Outcomes?

We can agree that the primary concern is with impacts of public spending on welfare. At one level, this seems uncontentious and straightforward. But in putting this objective into practice we quickly get into conceptual distinctions which may ultimately influence the judgements made about policy impacts. Although we may concur that human "welfare" is the objective, we may disagree about what that means.

Difficulties arise because there are multiple dimensions to welfare, and not all people adhere to the same concept. Views differ on the weight to assign different facets of well-being. Three general conceptualizations of welfare have been proposed: (i) *utility*, (ii) *income* and (iii) *capability*. Under the utility (or "welfarist") paradigm, individual utilities are the sole objective of policy.[1] Only individual preferences carry any weight and individual preferences are typically tied to both income and leisure. The "income" dimension is sometimes

interpreted as the monetary equivalent of utility where, for example, leisure is valued and added to other income. Alternatively, it is a distinct view where command over commodities is deemed to be the sole objective. More recently, the "capabilities" framework argues that welfare should be assessed by the attainment of certain basic capabilities, such as avoiding hunger and illiteracy (Sen 1985).

Three observations help illustrate why the multidimensionality of well-being, and the various views on what it means to be poor, complicate welfare assessments of policy impacts. First, even within a particular concept of welfare, there are likely to be important differences in how it is measured. For example, the methods of aggregating information on individual welfare impacts into a single summary statistic, such as an index of poverty can be contentious. How poverty is measured—the choice of the specific living standards indicator, the poverty cutoff point and the poverty index used—can all influence policy assessments (see Ravallion 1994). And even when there is agreement on the poverty measure, disagreements can still arise in how to assess policy options. For instance, some (Grosh 1995, Cornia and Stewart 1995 are examples) judge a policy by how well it concentrates benefits on the poor, so as to avoid errors in targeting (both errors in leakage to the non-poor and errors of imperfect coverage of the poor). Yet others, starting from the same objective of reducing income poverty, argue that the policy which has the greatest impact on poverty is not necessarily the one with the lowest errors of targeting—deliberate "errors" may well enhance the final impact on poverty by avoiding costs often associated with fine targeting (Ravallion and Datt 1995).

Second, it is rarely a case of choosing between one conceptualization or the other. Policy choice in practice is constrained by the availability of instruments and influenced by a

multiplicity of objectives. For example, interventions may target benefits exclusively to women for a variety of reasons. We may believe that women are more disadvantaged than men, or we may believe that there are positive externalities accruing to others as a result of targeting women (Appleton and Collier 1995). Similarly, as a way of raising incomes many support public spending on education because of its impact on productivity. The fact that education may also raise utility and capabilities—independently of incomes—is surely also worthy of consideration in our assessment of the policy against an alternative use of the funds.

A third observation is that while economists generally assume that utilities are the objective of policy (as well as of individual behavior), policymakers often focus more on non-utility concepts of welfare such as income—narrowly defined as command over commodities—or attainments in health and education. For example, policymakers concerned about reducing poverty do not typically consider the poor's leisure, or how hard they must work; even though there are costs to utility, they are not debited against gains in raising money income. Yet, so many of our rules-of-thumb derive from the utility framework. Certain consequences of this perspective for common targeting and evaluation rules are illustrated by Kanbur, Keen and Tuomala (1994). They show that if the policy objective is the minimization of an income-based poverty index, widely accepted rules-of-thumb based on the utility framework—namely that marginal tax rates on the poor should be low—are overturned. Under the alternative policy goal, simulated optimal marginal tax rates tend to be 60 percent or more (given minimal revenue requirements).

Another illustration is found in a study of the labor supply effects of Sri Lanka's food stamp scheme (Sahn and Alderman 1995). The study finds that the scheme resulted in both

men and women reducing their hours worked. Is this a good or bad outcome? That depends on one's welfare objective. If leisure is accorded a high weight, the policy may be considered a success. Parents may be spending more time with their children or they may be better off due to increased leisure time. But the policymaker trying to achieve the greatest dent in income poverty for a given budget may consider such behavioral responses costly. The same impact on income may have been achieved through a different policy at lower cost. Clearly, therefore, policy evaluation may depend on the underlying objective.

Consensus has not been reached on these choices. Indeed, disagreements are likely to persist. Given this state of affairs there is a need to recognize that there are multiple dimensions for judging policy impacts which cannot be easily aggregated into a widely acceptable measure. The message that underlying assumptions about welfare and poverty measurement influence the evaluation of public spending programs should be kept firmly in mind. This speaks to the need for clarity about those assumptions, and a recognition of how sensitive policy conclusions can be to changes in those assumptions.

How Can We Assess the Welfare Impacts of Public Spending?

Past endeavors to measure the welfare impact of public spending programs can be categorized into two general approaches. "Benefit incidence studies" and "behavioral approaches". The following discussion briefly describes each approach in order to draw out their strengths and limitations. Recent studies which blend the two approaches are then examined.

Benefit Incidence Studies

Benefit incidence studies assume that the value to consumers of a public service can be identified by the cost of providing it. They then assign benefits to the users of the service ranked by some agreed measure of current welfare. This provides a profile of the distribution of the specific category of public spending across the distribution of the chosen welfare indicator. There are various ways of presenting the results which allow one to determine whether the public spending component is progressive (inequality reducing) or regressive (inequality increasing); and whether it appears to be a good way to transfer benefits to the poor compared to other public spending components. An example is given by Figure 1 which shows how the benefits (in absolute monetary amounts) from spending on primary and tertiary education are distributed across income per capita groups ranked from poorest to richest for Tunisia in 1990 (Republic of Tunisia 1993). Results can be further dissagregated—for example across regional, gender and ethnic groups.

Benefit incidence studies divulge nothing about *why* incidence outcomes are what they are. For this reason the policy implications are limited and general, rather than specific. In reality, a number of factors conspire to produce the distributional outcomes. The variables determining the *supply* of, and the *demand* for, the public good or service, and how these vary across the distribution of living standards matter. For example, primary education is often found to be progressive because the poor (as defined by per capita income or consumption) tend to have more primary-school-age children. Further analysis of the underlying causes is often necessary in order to draw policy implications.

Benefit incidence studies have been around a long time. In the World Bank the twin studies by Meerman (1979) on Malaysia and by Selowsky (1979) on Colombia were very influential. Some governments also routinely carry out incidence studies, particularly in Latin America (e.g. Flood et al. 1995 for Argentina). Chile's household survey, CASEN, was designed with social sector benefit incidence analysis in mind.

The principle of benefit incidence studies is relatively straightforward. In practice, however, assembling the method's building blocks can be quite onerous. Simply ranking individuals correctly can be difficult. The data on utilization may be faulty, or not integrated with data on the welfare measure used. And piecing together the unit cost data—particularly in decentralized systems where funding may occur at various levels of government and there is no central accounting process—can be a nightmare.

On the positive side, the methodology is readily understood, the results easily presented and—especially when communicated graphically—they can be extremely powerful. The government may not have stopped to ponder the social welfare function implied by its spending policies. And even if it has, benefit incidence results may be a surprise to policy-makers. A benefit incidence study can be an influential tool in bringing home messages of the need for budget reallocations and reform.

Benefit incidence studies have had a great deal of influence in development policy. Many current policy recommendations are based on such studies. For example, the case for expanding the share of public spending on *basic* services—notably primary and secondary education and basic health care—is often based on benefit incidence studies. Such spending is found to reach the poor almost universally, while spending on tertiary services—university

education, hospital services—is invariably shown to be pro-rich by benefit incidence studies. (The education spending pattern exhibited in Figure 1 for Tunisia is typical of many countries.) Aggregate spending on basic health and education is invariably found to be inequality reducing though it may still confer lower absolute amounts to the poor than would a uniform (untargeted) transfer. Benefit incidence studies also find that certain food based schemes, social cash transfers, public employment schemes and other targeted transfer schemes have at times been quite pro-poor, though others have not, despite their pro-poor rhetoric.

SOME LIMITATIONS OF BENEFIT INCIDENCE STUDIES. Despite its influence, a number of concerns have been raised about the benefit incidence methodology. Some are serious enough to warrant considerable care in basing policy recommendations solely on evidence from such studies. Perhaps the most common criticism is that the unit cost of provision may have little relation to the value of the benefits to the individual. To take a simple example, the cost of immunizing a child is typically small compared to the lifelong benefits.

Another limitation is that benefit incidence studies give an incomplete representation of the welfare effects of public spending. This point relates back to the multi-dimensionality of well-being. Benefit incidence studies interpret all public spending—whether in cash or in-kind—in terms of the monetary transfer equivalent to see how those transfers alter the distribution of income or consumption.[2] However, we often want to know about impacts on other dimensions of living standards. For example, did health status improve as a result of health sector subsidies? Was the increased spending on schools reflected in higher literacy?

What was the impact of the school lunch program on nutrition and were there complementary effects on cognitive achievements? Policy makers are more often interested in how well a social sector policy performed in its intended effect (eg improved health status) than in how it topped off the distribution of income. One obvious implication of this line of criticism is that we should examine a range of impact measures and social indicators and not limit our analyses to benefit incidence.

A number of important public goods and services—important in terms of their likely impact on the poor—are not amenable to analysis by the benefit incidence methodology because it is unusually difficult to identify individual users or to estimate the unit cost of provision. Examples include safe water, sanitation, communicable disease and vector control and much physical infrastructure. All too often sector studies conduct benefit incidence analyses focusing solely on the spending for which users can be readily identified. Discussion of the rest of the government's budget—which may contain public spending components with the greatest consequence to the poor—falls through the cracks. This implies, once again, that benefit incidence studies need to be supplemented by other approaches.

Benefit incidence studies may not capture some important second-round effects on welfare. The method tries to identify direct transfer impacts. Yet, *indirect* benefits may be of considerable consequence to the distributional outcome. For example, while the poor are not direct beneficiaries of subsidies to tertiary education, the indirect benefits—transmitted through good governance, the existence of a class of technocrats, agricultural specialists, good educators, health personnel, and so on—may be of significance to the well-being and livelihood of the poor. Indirect benefits and externalities are difficult to estimate and so little is known

about their likely magnitude. But if they exist benefit incidence studies may be shaping policy recommendations which could be seriously flawed and have unfortunate consequences for the poor. This does not imply that governments should (to continue the same example) carry on subsidizing university education for the rich at the expense of even rudimentary primary schooling for the poor. Rather, it points to being more circumspect in drawing policy recommendations. For example, rather than public abandonment of universities in countries where there is no private market to take the slack and where markets, and credit markets in particular, do not work well, there may be room for alternative approaches. One possibility is to require a number of years of public service from graduates in repayment for free education benefits.[3] The government of Indonesia, for example, has long required the doctors it trains to spend some years working in remote rural primary health clinics upon graduation.

Benefit incidence studies present results on the distribution of *average* benefits. The *marginal* benefits distribution will often be of equal or greater interest in assessing public policy reform. Average incidence at one point in time may be a misleading indicator of the distribution of the *gains* from public spending (Lipton and Ravallion 1995, section 6.4.3). A seemingly beneficial expansion in the primary school budget may be buying better quality for schools in which the rich are enrolled rather than more public schools for the under-provisioned poor. One way to get at this is to replicate methodologies to compare incidence at two or more dates with comparable data to see how spending changes were distributed across different groups. Two recent studies use this technique and find that changes during the 1980s were pro-poor for Indonesia's public health sector and Malaysia's health and education sectors (van de Walle 1994, Hammer et al. 1995, respectively). Unfortunately the methodology is

unable to account for the factors underlying incidence patterns; for example, to what degree government policy as opposed to income growth can be credited with the improvements in equity. In order to draw policy lessons, we ultimately need to know what accounts for the favorable outcome.

The Malaysia study supplements the incidence analysis with more detailed analysis of the underlying mechanisms (Hammer et al. 1995). It attributes success in the education sector to the government's policy of ethnic targeting, but finds that pro-poor improvements in the health sector are due to the private sector's increasing ability to attract wealthier households. This study is a good example of the use of complementary methods to assess the distributional impacts of public spending policies.

Another promising avenue for exploring the marginal impacts of public expenditures is to use data which follow the same households or regions over time. The rising availability of such panel data and the econometric techniques for using them are paving the way for more sophisticated and robust estimation of distributional impacts and the introduction of dynamic effects in incidence. Using a panel of Hungarian households for 1987 through 1989, Ravallion et al. (1995) devise a methodology to examine how well the social safety net protected households from falling into poverty versus how well it promoted households out of poverty. Such evidence on dynamic performance would seem to be of key value in designing effective safety nets.

Perhaps the most damaging critique of benefit incidence studies is their assumption about the world without public spending—the counterfactual. The fundamental objective of a benefit incidence study is to compare the distribution of welfare with and without public

expenditures. Of course, we do not know what things would be like in the absence of public spending. So the without-intervention position is assumed to be the welfare indicator (e.g. income per capita) less the monetary value of the benefits secured from publicly-provided goods. However, from theory and overwhelming evidence, we know that public policies affect individual economic behavior, including labor supply, consumption, savings and investment decisions. A household's consumption pattern and its income depends on what public goods and services it expects to receive. We all make our work and savings decisions based on whether the government provides retirement pensions, whether we need to save for our children's education and whether we feel adequately protected in case of a medical emergency in the family. And, as is well recognized, such household responses have potentially important implications for a policy's final impact. Yet, benefit incidence studies simply ignore behavioral responses.

There has not been sufficient careful research to predict how much reranking of households would in fact occur if behavioral responses were factored in. The above-mentioned study of Hungary experiments with different estimates of the marginal propensity to consume out of social incomes to simulate the counterfactual and the consequences for assessments of the performance of the safety net (Ravallion et al. 1995). The results indicate that, although the distribution changes significantly, qualitative conclusions do not. Further research needs to test this in other settings and circumstances. In the absence of conclusive evidence on this issue, the validity of the conclusions drawn from benefit incidence studies must remain questionable.

A VERDICT ON BENEFIT INCIDENCE STUDIES? Where does the foregoing discussion of benefit incidence studies leave us? Some of the implications drawn from the studies—such as that governments should invest and reallocate budgets towards basic services—are compelling, partly because they reinforce tenets arising from economic theory and other evidence. Such services often fail to attract the private sector and thus accord with the principle that governments should be responsible for valuable goods which would otherwise be underprovided. Additional support comes from recent endogenous growth theory and a general consensus in the development community that human resource development is necessary for equitable economic growth (Bruno et al. (1996) review the theory and evidence). The fact that benefit incidence studies show basic services to be among the best ways to reach the poor surely clinches it.

Despite their limitations, benefit incidence studies can be useful in providing a rudimentary, first-cut approximation and working hypothesis of how specific components of public spending are distributed across key groups of interest. But, in actually using the results as endorsement for policy directions, much more circumspection is required than has tended to be the case. Before embarking on fundamental reform, governments should validate the policy implications with corroborating evidence from other sources. A broadening of the welfare indicator to cover non-income facets would also help. Benefit incidence results can be useful, but they do not provide sufficient information for charting the course of pro-poor reforms in public spending.

Behavioral Approaches

The second general approach to measurement, here denoted "behavioral approaches," can be interpreted as attempts to resolve the counterfactual problem described above by explicitly modelling behavioral responses. Also in contrast to benefit incidence studies, this class of methods is geared to estimating marginal incidence; and they can be used to explore the impacts of public spending on goods and services for which specific users cannot be identified, as well as to determine impacts on multiple dimensions of welfare.

Benefits are valued using either a monetary or non-monetary welfare metric consistent with behavioral responses to public spending. Compensating and equivalent variations and willingness-to-pay are examples of the monetary measures used in practice, while social indicators such as infant mortality or nutritional status are examples of non-monetary measures. By exploiting differences in policy across time or space, econometric techniques are used to estimate the effects of public spending programs or services on measures of well-being while controlling for other factors which could also be influencing outcomes. Household data are invariably needed to do this.

One lesson to come out of behavioral approaches is that behavioral responses are ubiquitous. Beneficiaries and non-beneficiaries adapt various aspects of their behavior to the existence of public intervention. A further finding has been that, although in the aggregate the demand for public services tends to be price inelastic, elasticities vary across income groups and the poor are invariably more price responsive. This has emerged from the literature on user fees (Gertler et al. 1987, Gertler and Glewwe 1989, Gertler and van der Gaag 1990). A single, static benefit incidence study may show (say) that the poor use public health clinics much more than other groups. But it will not reveal that the poor may also alter their usage

much more than others if prices increase. Benefit incidence studies ignore behavioral responses and may also overlook highly distributionally non-neutral behavioral effects.

SOME LIMITATIONS OF BEHAVIORAL APPROACHES. The behavioral approaches in general represent a more data intensive and technologically and methodologically complicated undertaking. Quite apart from these considerations, they have some potentially significant drawbacks. Here I focus on two particularly stubborn problems: biases in estimation and the problem of welfare identification.

BIASES IN ESTIMATION. Severe difficulties frequently arise in obtaining *unbiased* estimates of program effects. A fundamental problem in evaluation arises when the policy measure is not exogenous, but is correlated with the regression error term, so that resulting estimates of program impacts are biased.[4] There are various sources of such a correlation including simultaneity (whereby the policy's placement is determined in part by the measured welfare indicator) and omitted variable bias (whereby there is some omitted third variable which influences both program placement and the welfare outcome).[5] An example of the first would occur if a school feeding program were started in a village due to high undernutrition of children in that community. Nutritional status then both explains the existence of the intervention and is affected by it. Causality goes both ways. If, in contrast, the feeding program has ended up in the village as a result of the efforts of a local NGO that is taking an active interest in promoting the program and is raising well-being in the village independently of the program, then the second source of bias will be present. Unless the estimation allows

for the influence of the NGO, it will tend to confuse impacts due to the program itself with that due to the NGO's activities. This issue has long been recognized in policy evaluation, though its pervasiveness, the inadequacies of past corrective actions, and the sometimes dramatic consequences for policy recommendations have recently attracted renewed attention.[6]

The problem is common. In using cross-section household data to model policy impacts it can be generally presumed that there exist household or regional level fixed or time-varying effects which influence both policy enactment *and* policy outcomes. Endogeneity of program placement occurs when public services are located in response to local community or household factors not accounted for in the program's evaluation. Governments may locate programs in areas where they are desperately needed or where they are sure to have high returns. In this case, the omitted variable is likely to be a determinant of both the program variable and its outcome (the dependent variable). It is then impossible to sort out program impacts from the factors responsible for its placement. In a recent study, Pitt et al. (1993) implement a cross-regional fixed-effects methodology for dealing with the nonrandomness of program placement.[7] They show that the biases can be large. To take just one example, cross-sectional estimation, typically used in evaluation work, suggests that family planning facilities in Indonesia *increase* fertility. But once the authors follow an estimation method that allows for the fact that family planning facilities are deliberately set up in high fertility areas, this finding is reversed.

A second example in which this issue arises is when welfare impacts over time depend heavily on initial conditions through an endogenous growth process and those same initial conditions influence the location of interventions (Jalan and Ravallion 1996). Failure to measure fully the initial conditions at the time of intervention can then severely bias

assessments of future program benefits. In an attempt to evaluate the household level dynamic effects of a poverty alleviation program targeted to poor areas in China, Jalan and Ravallion (1996) find that community level factors (such as poor social and physical infrastructure) which ensure that the region is included in the target "poor areas" are also responsible for the lower rates of growth registered in those regions. The positive effects on living standards of the poor area development program can be entirely obscured if this dependence on initial conditions is not factored into the evaluation.

Besley and Case (1994) discuss the case where policy endogeneity is due to political and economic factors (for example changes in local political leadership and regional growth rates) which vary over time as well as space. In this case the fixed-effects methodology as used by Pitt et al. (1993) does not resolve the endogeneity problem.[8] Besley and Case illustrate how far wrong policy conclusions can be despite various common correction techniques. They explore the possibility of using political variables which influence policies but not welfare outcomes as a means of identifying policy impacts.[9]

One lesson from these studies is that the undetected biases in naive methods can be large—indeed, large enough to induce policymakers to abandon what are in effect successful policies, and even to pursue ones which a proper investigation would debunk.[10] Panel data can go some way toward dealing with these problems, such as through fixed effects regression models (Pitt et al. 1993) and dynamic models of household level consumption growth allowing for state dependence (Jalan and Ravallion 1996). Instrumental variable techniques can also be useful in some cases. An understanding of government budget allocation and program placement rules, and of political economy more generally, may be fundamental to coming to

grips with public spending impacts (Besley and Case 1994, Pitt et al. 1993). This requires collection of new sorts of data (such as time series of the "history" of government policy and of political changes) and imaginative ways of getting at—and incorporating into measurement methodologies—the institutional contexts within which households and governments make decisions. This underlines the need to collect and link various sources and levels of data: household, community, government local public finance and environmental. It is a good idea to begin by designing new surveys so that they fit the institutional features of local public finance—including the entire tax and expenditure system—ensuring that the two complementary sources of data can be meshed together. Doing so will allow data and research to better inform public action.

Finally, using randomized control (or experimental) designs for evaluating certain policy interventions also resolves biases due to endogeneity problems (discussions are given in Grossman 1994, Newman et al. 1994, and Kremer 1995). However, randomly assigning policy interventions is not always feasible. Often too, we will also be interested in the model which generates outcomes. Random assignment methods are not useful in this case since they only provide estimates of mean outcomes with and without the intervention and reveal little about the underlying factors which determine those outcomes.

IDENTIFYING WELFARE. Our information on individual and household behavior typically comes from household surveys with information on household characteristics and the consumption of market goods. Common practice in implementing utility-based measures of welfare is to use the survey information to reveal preferences and get at the underlying

"behaviorally consistent" indicator of well-being. For example, one approach to setting equivalence scales (expenditure levels needed by households of different sizes and demographic compositions to achieve the same utility) starts by estimating a model of consumer demand which is used to infer the equivalence scale. Such welfare measures have limitations, particularly when consumption includes publicly-provided public and private goods. It is now well recognized that demand behavior for market goods does not provide sufficient information for identifying utility when—as is invariably the case—there are non-market goods which matter to welfare (see Pollak 1991, and Browning 1992). For this reason, attempts at constructing welfare measures are often inadequate or contentious. It is important to understand the limitations of welfare measures which we calculate in this way.

This problem could be avoided if the welfare indicator were directly observable. Many studies focus on measurable health and educational outcomes which implicitly have direct welfare significance. They econometrically estimate a reduced form relationship linking the welfare indicator to inputs, including socioeconomic characteristics of regions and households, utilization and availability of public and private services. Unfortunately, a direct but *partial* indicator of welfare does not solve the identification problem. Because such approaches are not grounded within a broader welfare framework that would allow a valuation of the benefits from an improved outcome or a public investment, they have unclear welfare and policy interpretations. By this approach we learn the impact of a given investment on both health and education outcomes (say). In both cases public spending may be having considerable impact but we are left without a means to determine the appropriate tradeoffs between them. Which should receive priority?

Such complications in valuation can be attributed to two fundamental problems (Cornes 1995). The first is that in the real world, prices and other individual or household characteristics vary across individuals. The same expenditure level will entail a lower level of welfare for a large household living in a remote rural area where most prices are high than for a small household living in a low price region. Secondly, quantity constraints are common; individuals are often forced to consume more or less than they would like because goods are lumpy (indivisible) or because there is rationing of some sort. In both cases, the available quantity is unlikely to equal the desired quantity. The price paid may not then reflect the true value to the consumer. Public goods are intrinsically lumpy. Conventional methods assume that each consumer can purchase anything she wants at prevailing prices and budget constraints. But under rationing, conventional measures of real incomes are likely to be imperfect measures of welfare. For example, a household's "real" (price and demographics-adjusted) income may allow it to buy the 2 breads per day it would like, but if the store rations the household to one bread only, then it will be worse off than another household with the same real income which can buy as much as it wants. In such a case, comparisons of pre- and post-intervention incomes or consumption expenditures adjusted for prices may distort the picture of what has actually happened to welfare.

New Approaches Incorporate Behavioral Responses into Incidence Studies.

So far, benefit incidence studies and behavioral approaches have been discussed separately. Why not combine the approaches? A number of recent studies can be interpreted

as moving in that direction. These "*hybrid*" approaches do not necessarily solve all the problems but they can certainly help.

By measuring benefits net of behavioral responses, the behavioral approach tries to determine the true impact of policy. It often does this in the aggregate for some "representative household". Regressions predict mean outcomes. The next step is to determine impacts at a more dissaggregated level and to assign those impacts correctly in the distribution of welfare. In other words, the aim is to make policy impacts a *function* of household characteristics, including income, not just control for them. If this can be done then the behavioral approach can also reveal the incidence of benefits.

A series of recent studies illustrate how econometrically estimated parameters and simulation techniques allow what are basically benefit incidence studies to be modified so as to incorporate incentive effects and better estimate the distribution of a policy's net benefits across households. None of these studies attempts to look at all the potential behavioral effects or second-round effects, focusing instead on a specific parameter which is deemed to be the most crucial behavioral response in the particular policy circumstances. Thus, although these efforts represent an important first step in understanding how behavioral responses may alter a policy's fiscal incidence, their results cannot be considered as providing conclusive evidence on the specific policies examined.

Labor supply responses to interventions are often a concern for policymakers. The poor care about their leisure. However, if a policy is aiming to maximize income gains to the poor for a given budget, labor supply responses are going to be an important factor to look at. One recent study examines how the labor supply of beneficiaries responded to Sri Lanka's food

stamp scheme and the implications for the scheme's net transfers (Sahn and Alderman 1995). Labor supply is modelled and the parameters used to simulate the counterfactual of what labor market effort, and hence incomes, would have been had the food stamp program not existed. Receiving the food-stamp is estimated to have reduced work by as much as 3 days per month for men and women in rural areas corresponding to around 30 percent of the gross transfer from the scheme.

It has been claimed that the behavior of *non*beneficiaries may also influence a policy's net impact. For example, in many societies and, in particular, in poor developing countries there is an active "moral economy" among subgroups of households; group members share risks and redistribute in favor of the poor. But households who are helping less well-off relatives or other group members may stop doing so if the government introduces schemes to aid households. Cox and Jimenez (1995) look at this issue and provide another example of an attempt to assess how behavioral responses may alter a policy's final distributional impact. The determinants of net inter-household transfers received in the absence of public intervention in the Philippines are estimated and predicted parameters used to simulate the likely private transfer response to government redistributive programs. The results suggest that public provision of social security may displace the private moral economy that exists between households in the Philippines.

A study of a public employment program aimed at reducing poverty in India attempts to estimate the income forgone by participants in order to participate (Datt and Ravallion 1994; Ravallion and Datt 1995). This is a cost of participation which must clearly be netted out from the benefits to get at the actual impact of the scheme on poverty. But this is rarely done. The

study aims to estimate the counterfactual of what participants would have been doing had they not been working on the public employment program. Household level data for two villages are used to estimate a time allocation model which predicts the time the household devoted to various activities as a function of the employment scheme and other exogenous variables. Tests are made for exogeneity of public employment. This model is then used to simulate the time displacement due to the scheme. The income losses are valued to see what the cost of participation was and how its distribution alters the scheme's impact. The costs are found to be around 25 percent of gross wage earnings from the scheme.

A somewhat different example is given by a study which explores how the gains from an expansion of irrigation infrastructure in Viet Nam are likely to be distributed given household and region specific characteristics (van de Walle 1996a, 1996b). Crop incomes are modelled with special emphasis on the marginal impacts of irrigated and non-irrigated land for different types of households in different regions. The results are then used to simulate the potential size and cross-household distribution of gains from investments in irrigation. The marginal effects of irrigation on crop incomes are found to vary with household demographics, education and region of residence. Together with the existing distribution of irrigated and non-irrigated land across households, these factors are shown to have significant influence on the incidence of benefits and the distributional impacts of a program of irrigation expansion.[11] [12]

These studies demonstrate that it is feasible to incorporate behavioral responses into the analysis of the incidence of public spending. They also indicate that the effects can be quite pronounced. There is bound to be much more empirical research of this type in the near future. In addition to the important caveats concerning policy endogeneity and the

identification of true welfare effects, another warning concerns the use of simulations. In all the studies discussed, simulation techniques are used to draw out behavioral implications. These may dictate that new assumptions be introduced. Naturally, care must be taken that such assumptions do not in turn feed into the results and ultimately the conclusions from the studies.

BEHAVIORAL RESPONSES IN SUM. There is incontrovertible evidence that the behavior of beneficiaries and non-beneficiaries responds in all sorts of ways to public interventions. But there is a need to understand *when* ignoring behavioral responses will matter most. It should not automatically be assumed that more complicated models will produce better or fundamentally different policy advice. Rather little is known about how much difference behavioral responses can make to key policy conclusions. At a quite general level, more research is needed to determine whether they do or not, so as to guide "rules of thumb" found in practice.

In a specific country, it will clearly not be feasible to launch careful, rigorous, impact evaluations and complicated econometric or randomized assignment methods for each new public policy and program initiative or reform. A partial reliance on "quick and dirty" methods, such as benefit incidence studies, is likely to be necessary for quite some time. Yet, while we remain substantially ignorant about the policy implications of behavioral responses and other second round effects, periodic investigations of the more full blown behavioral and other effects can have high payoffs. These methods can provide an important reality check and contribute to a much richer and honest perspective on the consequences of policy changes for various aspects of living standards.

Conclusions

Public spending policies are continually being enacted and reformed based on assessments of their effects on poverty and inequality in living standards. This paper has surveyed the methods most typically used to evaluate the distributional impacts of public expenditures. In representing the current state of the art, the paper has attempted to identify some of its limitations and draw implications for best practice.

Evaluating a policy's impact requires assessing how different things would have been in its absence. However, the counterfactual of no intervention is often tricky to quantify precisely. Benefit incidence studies ignore behavioral responses and second round effects, and simply use the cost of provision as a proxy for benefits received. Behavioral approaches present quite different drawbacks in their attempts at representing individual benefits correctly. Recent studies attempt to incorporate behavioral responses into incidence assessments. However, uncertainty remains about whether behaviorally consistent methods actually point to fundamentally different policy recommendations. What can be concluded is that we need to diversify and compare results from our evaluation methods and broaden our definition of well being to see how various facets of living standards are affected.

Notes

1. On the "welfarist" versus "non-welfarist" distinction, see Sen 1979.

2. In terms of the three conceptualizations of welfare discussed earlier, benefit incidence analysis is squarely focused on the income dimension alone.

3. Such a policy can be interpreted as a way of making credit markets work better.

4. To obtain unbiased estimates of the parameters in a regression model—giving the outcome variables as a function of various explanatory variables—one requires that the latter variables are not correlated with the regression's error term, embodying measurement error in the dependant variable and omitted variables.

5. Pointing to its many possible sources, Deaton (1995 pp 1832-33) summarizes the problem thus: "...when we want to estimate the effects of a policy or a project, we must take into account what determines it, and having done so, we will usually find that we cannot discover its effects by standard regressions. The basic issue here is the correlation of explanatory variables with the error term, and it matters less whether we think of that correlation as coming from simultaneity, heterogeneity, selection, or omitted variables." In all these cases, the exogeneity assumption does not hold.

6. Meyer (1995) and Besley and Case (1994) discuss these issues and scrutinize some of the recent policy evaluation analysis for developed countries. Pitt et al. (1993), Jalan and Ravallion (1996), McKernan (1996), and Pitt and Khandker (1996) are examples of recent studies which discuss and attempt to deal with biases in estimation of policy impacts due to various forms of endogeneity in developing country applications. Strauss and Thomas (1995) review attempts to circumvent such biases in the context of estimating program effects on health and education outcomes in poor countries.

7. The fixed-effects technique exploits the availability of multiple observations over time to rid variables of any time-invariant (fixed) characteristics which influence both policies and their outcomes.

8. The fixed-effects methodology may have other drawbacks. For example, if variation over time contains a lot of measurement error, eliminating the fixed-effects creates another problem—namely that changes over time are now heavily contaminated by measurement error (including in explanatory variables) thus biasing estimates (Deaton 1995).

9. The political variables then act as instruments for the policies; in the regression of welfare outcomes on policies, the actual policies are then replaced by their predicted values obtained by first regressing on the political variables.

10. Besley and Case (1994) illustrate in the context of workers' compensation benefits. The debate between Hanushek (1995) and Kremer (1995) on estimations of the effectiveness of

education spending, impacts on outcomes, and what policy conclusions to draw, provides another interesting example.

11. For example, marginal gains from irrigated land are found to be higher for educated households. The fact that education tends to be higher in (poorer) North Viet Nam as well as more equitably distributed there, influences the progressivity of benefit incidence (van de Walle 1996b). A conventional benefit incidence study would have typically assumed marginal benefits to be constant and hence, would have underestimated the potential progressivity of irrigation expansion.

12. Note that the income gains from irrigation are unlikely to be reaped currently. Cross-sections survey households in different long run situations. That variation is used to infer effects of a policy which may take a long time to occur. Irrigation benefits may not be immediate but the methodology cannot reveal when they will occur. This is a general problem in interpreting the implications from static models.

References

Appleton, Simon and Paul Collier. 1995. "On Gender Targeting of Public Transfers". In van de Walle and Nead (1995).

Besley, Timothy and Anne Case. 1994. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies". NBER Working Paper 4956, NBER, Cambridge, Ma.

Browning, Martin. 1992. "Children and Household Behavior." *Journal of Economic Literature* 30:1434-75.

Bruno, Michael, Lyn Squire and Martin Ravallion. 1996. "Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues." Policy Research Working Paper 1563, World Bank, Washington, D.C.

Cornes, Richard. 1995. "Measuring the Distributional Impact of Public Goods". In van de Walle and Nead (1995).

Cornia, Giovanni Andrea and Frances Stewart. 1995. "Two Errors of Targeting". In van de Walle and Nead (1995).

Cox, Donald and Emmanuel Jimenez. 1995. "Private Transfers and the Effectiveness of Public Income Redistribution in the Philippines". In van de Walle and Nead (1995).

Datt, Gaurav and Martin Ravallion. 1994. "Transfer Benefits from Public Works Employment: Evidence from Two Indian Villages" *The Economic Journal* 104: 1346-69.

Deaton, Angus. 1995. "Data and Econometric Tools for Development Analysis" In Behrman, Jere and T.N. Srinivasan (eds.) *Handbook of Development Economics Volume 3*. Amsterdam: North-Holland.

Flood, M. Cristina, Marcela Harriague, Leonardo Gasparini and Benigno Velez. 1994. *El Gasto Publico Social Y Su Impacto Redistributivo*. Secretaria de Programacion Economica, Buenos Aires.

Gertler, Paul and Paul Glewwe. 1989. "The Willingness to Pay for Education in Developing Countries." Living Standards Measurement Study Working Paper No. 54, Washington, D.C.

Gertler, Paul, Luis Locay and Warren Sanderson. 1987. "Are User Fees Regressive?: The Welfare Implications of Health Care Financing Proposals in Peru." *Journal of Econometrics* 33:67-88.

Gertler, Paul and Jacques van der Gaag. 1990. *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries*. Baltimore: The Johns Hopkins University Press.

Grosh, Margaret E. 1995. "Towards Quantifying the Tradeoff: Administrative Costs and Incidence in Targeted Programs in Latin America". In van de Walle and Nead (1995).

Grossman, Jean Baldwin, 1994. "Evaluating Social Policies: Principles and U.S. Experience," *The World Bank Research Observer* 9(2): 159-180.

Hammer, Jeffrey, Ijaz Nabi, and James Cercone. 1995. "Distributional Impact of Social Sector Expenditures in Malaysia". In van de Walle and Nead (1995).

Hanushek, Eric. 1995. "Interpreting Recent Research on Schooling in Developing Countries". *The World Bank Research Observer* 10(2): 227-246.

Jalan, Jyotsna and Martin Ravallion. 1996. "Are there Dynamic Gains from a Poor-Area Development Program?," PRDPH, unprocessed, World Bank.

Kanbur, Ravi, Michael Keen, and Matti Tuomala. 1994. "Labor Supply and Targeting in Poverty Alleviation Programs". *World Bank Economic Review* 8(2): 191-211.

Kremer, Michael. 1995. "Research on Schooling: What We Know and What We Don't: A Comment on Hanushek". *The World Bank Research Observer* 10(2): 247-254.

Lipton, Michael and Martin Ravallion. 1995. "Poverty and Policy." In Behrman, Jere and T.N. Srinivasan (eds.) *Handbook of Development Economics Volume 3*. Amsterdam: North-Holland.

McKernan, Signe-Mary. 1996. "The Impact of Micro-Credit Programs on Self-Employment Profits: Do Non-Credit Program Aspects Matter?" unprocessed, Department of Economics, Brown University.

Meerman, Jacob. 1979. *Public Expenditure in Malaysia: Who Benefits and Why*. New York: Oxford University Press for the World Bank.

Meyer, Bruce. 1995. "Natural and Quasi-Experiments in Economics". *Journal of Business and Economic Statistics* April: 151-161.

Newman, John, Laura Rawlings, and Paul Gertler, 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries," *The World Bank Research Observer*. 9 (2): 159-180.

Pitt, Mark M., Mark R. Rosenzweig, and Donna M. Gibbons. 1993. "The Determinants and Consequences of the Placement of Government Programs in Indonesia". *World Bank Economic Review* 7(3): 319-48.

Pitt, Mark M. and Shahidur Khandker. 1996. Household and Intrahousehold Impacts of the Grameen Bank and Similar Targeted Credit Programs in Bangladesh. World Bank Discussion Paper 320. World Bank, Washington, D.C.

Pollak, Robert A.. 1991. "Welfare Comparisons and Situation Comparisons." *Journal of Econometrics* 50:31-48.

Ravallion, Martin. 1994. *Poverty Comparisons*. Fundamentals in Applied Economics Volume 56, New York: Harwood Academic Press.

Ravallion, Martin and Gaurav Datt. 1995. "Is Targeting Through a Work Requirement Efficient? Some Evidence from Rural India." In van de Walle and Nead (1995).

Ravallion, Martin, Dominique van de Walle, and Madhur Gautam. 1995. "Testing a Social Safety Net" *Journal of Public Economics*, 57:175-199.

Republic of Tunisia. 1993. "Les Dépenses Budgetaires de L'Éducation et de L'Enseignement Superieur: Répartition selon les Groupes Sociaux et Évolution entre 1985 et 1990." Institut National de la Statistique, Tunis.

Sahn, David E. and Harold Alderman. 1995. "Incentive Effects on Labor Supply of Sri Lanka's Rice Subsidy". In van de Walle and Nead (1995).

Selowsky, Marcello. 1979. *Who Benefits from Government Expenditures? A Case Study of Colombia*. New York: Oxford University Press.

Sen, Amartya. 1979. "Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics?" *The Economic Journal* 89:537-58.

Sen, Amartya. 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.

Strauss, John and Duncan Thomas. 1995. "Human Resources: Empirical Modeling of Household and Family Decisions." In Behrman, Jere and T.N. Srinivasan (eds.) *Handbook of Development Economics Volume 3*. Amsterdam: North-Holland.

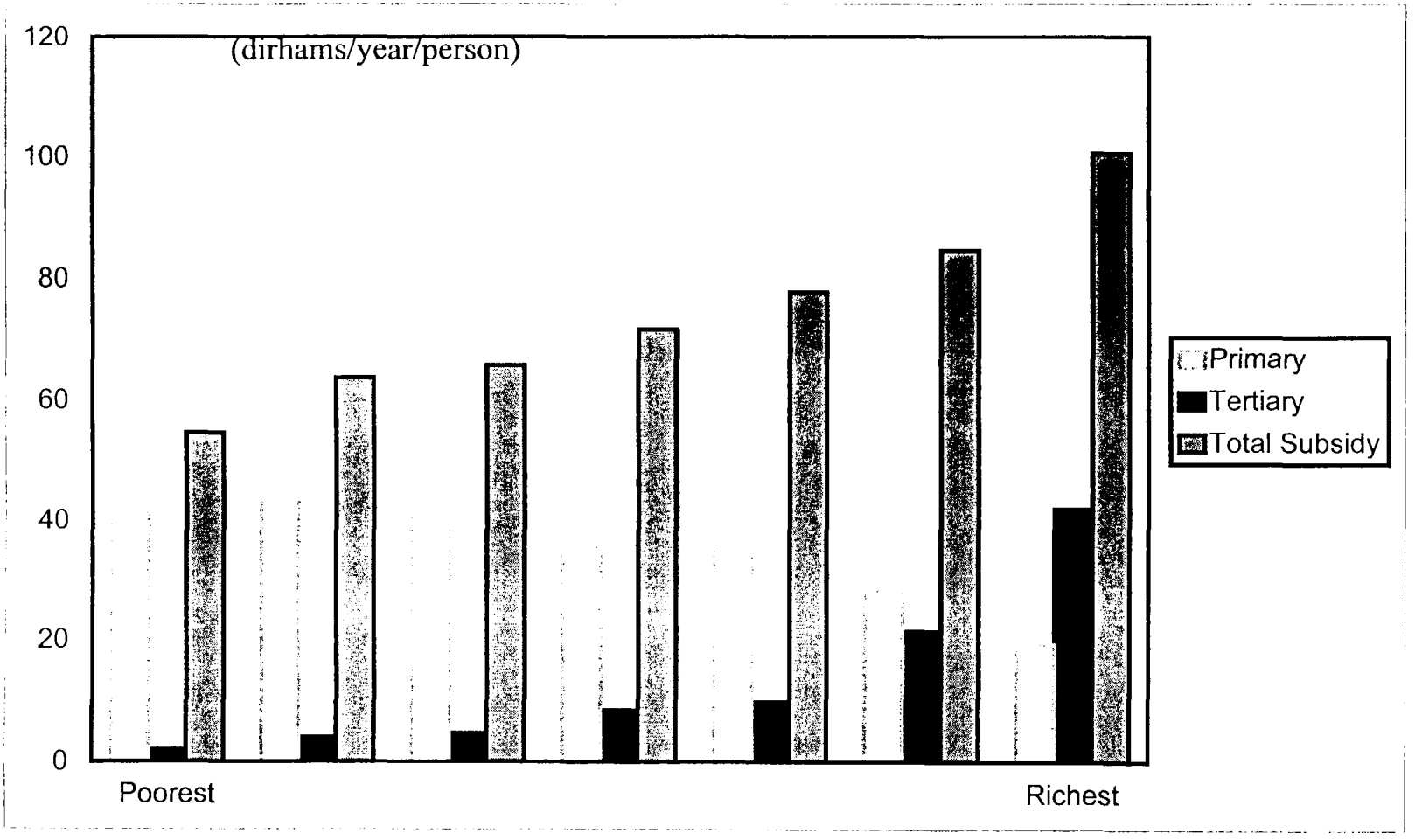
van de Walle, Dominique and Kimberly Nead. 1995. *Public Spending and the Poor: Theory and Evidence*. Baltimore and London: The Johns Hopkins University Press.

van de Walle, Dominique. 1994. "The Distribution of Subsidies Through Public Health Services in Indonesia 1978-87". *World Bank Economic Review* 8(2): 279-309.

van de Walle, Dominique. 1996a. "Infrastructure and Poverty in Viet Nam". Living Standards Measurement Study Working Paper No. 121, Washington, D.C.

van de Walle, Dominique. 1996b. "Are there Complementarities Between Human and Physical Infrastructure in Developing Countries? A Case Study for Irrigation in Viet Nam," unprocessed, Policy Research Department, World Bank, Washington, D.C.

Figure 1: Subsidies to Primary and Tertiary Education, Tunisia, 1990



Source: Republic of Tunisia, 1993.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1651	Bank-Led Restructuring in Poland: Bankruptcy and Its Alternatives	Cheryl W. Gray Arnold Holle	September 1996	B. Moore 38526
WPS1652	Intra-Industry Trade, Foreign Direct Investment, and the Reorientation of Eastern European Exports	Bernard Hoekman Simeon Djankov	September 1996	F. Hatab 35853
WPS1653	Grants and Debt Forgiveness in Africa: A Descriptive Analysis	Leonardo Hernández	September 1996	R. Vo 31047
WPS1654	Indonesia's Palm Oil Subsector	Donald F. Larson	September 1996	P. Kokila 33716
WPS1655	Uncertainty and the Price for Crude Oil Reserves	Timothy J. Considine Donald F. Larson	September 1996	P. Kokila 33716
WPS1656	The Investment Decision: A Re-Examination of Competing Theories Using Panel Data	Cherian Samuel	September 1996	C. Samuel 30802
WPS1657	Is There an Optimal Structure for Decentralized Provision of Roads?	Frannie Humplick Azadeh Moini-Araghi	September 1996	J. Williams 82557
WPS1658	Decentralizing Structures for Providing Roads: A Cross-Country Comparison	Frannie Humplick Azadeh Moini-Araghi	September 1996	J. Williams 82557
WPS1659	Unemployment Insurance in Algeria: Implications for a Labor Market in Transition	Elizabeth Ruppert	September 1996	H. Osselyn 36039
WPS1660	Mind Your P's and Q's: The Cost of Public Investment is *Not* the Value of Public Capital	Lant Pritchett	October 1996	S. Fallon 38009
WPS1661	Determinants of Public Expenditure on Infrastructure: Transportation and Communication	Susan Randolph Zeljko Bogetic Dennis Heffley	October 1996	Z. Bogetic 623-7292
WPS1662	From Learning to Partnership: Multinational Research and Development Cooperation in Developing Countries	Giorgio Barba Navaretti Carlo Carraro	October 1996	M. Patena 39515
WPS1663	Internal Finance and Investment: Another Look	Cherian Samuel	October 1996	C. Samuel 30802
WPS1664	Pensions in Germany	Monika Queisser	October 1996	H. Arbi 34663

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1665	How Important Are Labor Markets to the Welfare of Indonesia's Poor?	Andrew D. Mason Jacqueline Baptist	October 1996	D. Ballantyne 87198
WPS1666	Is Growth in Bangladesh's Rice Production Sustainable?	John Baffes Madhur Gautam	October 1996	P. Kokila 33716
WPS1667	Dealing with Commodity Price Uncertainty	Panos Varangis Don Larson	October 1996	J. Jacobson 33710
WPS1668	Small is Beautiful: Preferential Trade Agreements and the Impact of Country Size, Market Share, Efficiency, and Trade Policy	Maurice Schiff	October 1996	M. Patena 39515
WPS1669	International Capital Flows: Do Short-Term Investment and Direct Investment Differ?	Punam Chuhan Gabriel Perez-Quiros Helen Popper	October 1996	T. Nadora 33925
WPS1670	Assessing the Welfare Impacts of Public Spending	Dominique van de Walle	October 1996	C. Bernardo 31148